Why Bostrom's Argument Is Ill-Formed: The Problem of the Missing Subject

A deconstruction of the simulation trilemma through analysis of its hidden epistemological premise

Nick Bostrom's trilemma, put forward in 2003, remains one of the most widely discussed philosophical arguments of our time. Let me briefly restate it. One of three propositions must be true. Either civilizations generally perish before they acquire the ability to create full-fledged simulations of consciousness. Or those that survive choose, for whatever reason, not to do so. Or we are, in all likelihood, already living inside such a simulation.

For twenty years philosophers, physicists, and specialists in artificial intelligence have debated which option is the most probable. Hundreds of papers have been written. Probabilistic models, counterarguments, and refinements have been proposed. Yet amid all this variety of positions, there is one thing that, as far as I am aware, no one has called into question: **the very correctness of the question itself**.

I want to show that the question is ill-formed. Not because Bostrom's logic is flawed — it is impeccable. But because at the foundation of the argument lies a premise that no one has examined.

Who Decides?

Look at the formulations carefully. Civilization goes extinct. Civilization chooses not to create simulations. Civilization builds simulations in which we live. In all three cases the actor is the same — civilization. It functions as a subject, an agent that makes decisions or suffers consequences.

But what is civilization as a subject? This is not a metaphor or a convention. For the trilemma to work, civilization must be capable of something like collective choice. Not voting, not market equilibrium, but genuine choice — a conscious act concerning its own nature.

This is where a question arises that I consider crucial and that, it would seem, has not yet been posed in explicit form: **does such a subject exist?**

Five Steps Toward an Answer

I propose the following chain of reasoning. Each step in it, as I see it, can be verified independently.

First. Bostrom's argument requires a civilizational subject. Without an agent capable of going extinct, refusing, or creating, the three options of the trilemma lose their meaning. This is not an interpretation — it is a direct consequence of the argument's structure.

Second. If civilization is a subject, it must possess something like an organ of self-knowledge. A subject that is incapable of reflecting on its own nature cannot "decide" whether or not to create simulations. That would be tantamount to attributing the capacity for choice to a system that has no mechanism for recognizing alternatives.

Third. No such organ exists. I state this as a verifiable fact and invite any opponent to refute it. As of today, there is not a single institution, not a single body, not a single interdisciplinary program that is authorized to investigate the question of the nature of civilization as a whole — not within philosophy, not within physics, not within computer science, but precisely as an integral problem requiring the simultaneous consideration of all these perspectives. Bostrom's question belongs to no discipline. It falls into the space between them — where there are no departments, no journals, no experts.

I call this space the Central Corridor of Recognition. This is neither a theory nor a metaphor. It is an observation: there exist questions that are visible from the space between disciplines but for which, inside the disciplinary system, there is no addressee. Bostrom's question is one of them.

Fourth. If no organ of self-knowledge exists, then civilization is not a subject. It is an environment. An aggregate of individuals, each of whom thinks within the framework of their own discipline, their own experience, their own neurological limitations, does not constitute an agent. Neurology is unambiguous here: the human brain is organized around individual perception. Collective consciousness is an appealing metaphor, but it has no neurological substrate. We are not ants and not cells of a single organism. We are separate people in separate rooms.

Fifth. If the subject does not exist, Bostrom's argument loses its addressee. The trilemma describes the choice of an agent that is not there. It is neither false nor true. It is ill-formed — in the same sense in which the question "what color is envy?" is ill-formed. Grammatically flawless, logically empty.

What Can Be Verified

The central claim of this paper — the absence of an institution capable of investigating the question of the nature of civilization as a whole — is open to direct verification. It is enough to try to find such an institution.

No university in the world has a department whose task is to investigate the question "is our civilization a simulation?" as an integral problem rather than as a topic for a seminar in philosophy of mind. No government body is authorized to pose this question. No grant program funds its investigation in a holistic formulation.

This is not an oversight or an omission. I believe it to be a structural property of the existing system of knowledge. The system is organized in such a way that integral questions are inevitably broken into disciplinary fragments. Each fragment is investigated conscientiously, but the whole is assembled nowhere and by no one. The corridor between the rooms is empty.

Where This Leads

I do not claim that we are living in a simulation. I do not claim the opposite. I claim something more modest and, it seems to me, more important: **we do not have a subject on whose behalf this question could be posed**.

Twenty years of discussion around Bostrom's trilemma have been, in essence, a discussion within an ill-formed question. Everyone who chose one of the three options implicitly accepted the existence of a civilizational subject — and thereby reproduced the error embedded in the foundation of the argument.

This observation, in turn, raises its own question: why has no one noticed the missing subject in twenty years? Perhaps the answer is that doing so requires standing not inside any one discipline but in the space between them. In a corridor from which all the rooms are visible, but which itself belongs to no one.

This work was carried out using the CCR method (Central Corridor of Recognition) — an approach that investigates not the content of disciplinary answers but the conditions under which certain questions can or cannot be posed.